The Annals of Statistics
2012, Vol. 40, No. 1, 159–187
DOI: 10.1214/11-AOS954

LARGE-SAMPLE STUDY OF THE KERNEL DENSITY ESTIMATORS UNDER MULTIPLICATIVE CENSORING[1]

By Masoud Asgharian[2], Marco Carone[3] and Vahid Fakoor

McGill University, University of California, Berkeley, and Ferdowsi University of Mashhad

The multiplicative censoring model introduced in Vardi [*Biometrika* **76** (1989) 751–761] is an incomplete data problem whereby two independent samples from the lifetime distribution G, $\mathcal{X}_m = (X_1, \ldots, X_m)$ and $\mathcal{Z}_n = (Z_1, \ldots, Z_n)$, are observed subject to a form of coarsening. Specifically, sample $\mathcal{X}_m$ is fully observed while $\mathcal{Y}_n = (Y_1, \ldots, Y_n)$ is observed instead of $\mathcal{Z}_n$, where $Y_i = U_i Z_i$ and $(U_1, \ldots, U_n)$ is an independent sample from the standard uniform distribution. Vardi [*Biometrika* **76** (1989) 751–761] showed that this model unifies several important statistical problems, such as the deconvolution of an exponential random variable, estimation under a decreasing density constraint and an estimation problem in renewal processes. In this paper, we establish the large-sample properties of kernel density estimators under the multiplicative censoring model. We first construct a strong approximation for the process $\sqrt{k}(\hat{G} - G)$, where $\hat{G}$ is a solution of the nonparametric score equation based on $(\mathcal{X}_m, \mathcal{Y}_n)$, and $k = m + n$ is the total sample size. Using this strong approximation and a result on the global modulus of continuity, we establish conditions for the strong uniform consistency of kernel density estimators. We also make use of this strong approximation to study the weak convergence and integrated squared error properties of these estimators. We conclude by extending our results to the setting of length-biased sampling.

1. Introduction. Vardi [50] introduced an incomplete data problem unifying several statistical models. The problem consisted of inferring the lifetime distribution of interest G through a random sample $X_1, X_2, \ldots, X_m$

Received May 2009; revised July 2011.

[1] Authorship is listed alphabetically to reflect the equal contribution of each author.

[2] Supported in part by NSERC, FQRNT and the Veterans Affairs HSR&D Service.

[3] Supported in part by NSERC, FQRNT, NIH and the Hopkins Sommer Scholars Program.

AMS 2000 subject classifications. Primary 62N01; secondary 62G07.

Key words and phrases. Integrated squared error, kernel density estimation, length-biased sampling, modulus of continuity, multiplicative censoring, strong approximation.

drawn directly from G and a random sample $Y_1, Y_2, \ldots, Y_n$ drawn from the distribution F with density function

$$f(y) = \int_{y \leq z} z^{-1}\, dG(z), \qquad y > 0. \tag{1.1}$$

Since f is a decreasing density function, Y may be expressed as the product of two independent random variables: a nonnegative variate Z and a standard uniform variate U. From the form of (1.1), it is easy to see that in this case Z must be distributed according to G. This representation suggests that only a random fraction of Z may be observed, motivating the nomenclature *multiplicative censoring* used to describe this incomplete data scheme. The likelihood based on the $k = m + n$ observations $X_1 = x_1, \ldots, X_m = x_m$ and $Y_1 = y_1, \ldots, Y_n = y_n$ is

$$L(G) = \prod_{i=1}^{m} G(dx_i) \prod_{j=1}^{n} \int_{y_j \leq z} z^{-1}\, dG(z). \tag{1.2}$$

As discussed by Vardi [50], the multiplicative censoring model arises from the deconvolution of an exponential random variable, estimation under a decreasing density constraint and an estimation problem in renewal processes. The literature on these and related problems is vast. Estimation under a decreasing density constraint dates back to the seminal work of Grenander [22], with key contributions by Groeneboom [23] and Huang and Wellner [26]. The estimation problem in renewal processes discussed in [50] is closely tied to important applications in cross-sectional sampling and prevalent cohort studies in epidemiology (length-biased sampling) and in labor force studies in economics (stock sampling). The multiplicative censoring model and its variants have been studied by [6, 8, 25, 45, 50] and [51], among others. Vardi [51] studied the asymptotic behavior of solutions of the nonparametric score equation under the multiplicative censoring model.

As will be discussed later, multiplicative censoring and left-truncated right-censored data are intricately tied. The latter have been extensively studied in the statistical literature. Their importance stems mainly, although not exclusively, from the widespread use of prevalent cohort study designs to estimate survival from onset of a disease. In such studies, patients with prevalent disease are identified at some instant in calendar time, often through a cross-sectional survey. These patients are then followed forward in time until death or loss to follow-up. If no temporal change in the incidence of disease has occurred during the period covering observed onsets, a stationary Poisson process may adequately describe the incidence pattern of the disease; see [2–4] and [53]. In this case, the left-truncation variable is uniformly distributed, and the failure time data are said to be length-biased.

The likelihood for the observed data is then given by (1.2), where

$$G(t) = \mu_U^{-1} \int_0^t u \, dF_U(u),$$

$\mu_U = \int_0^\infty u \, dF_U(u)$ and F_U, the *unbiased distribution*, is the underlying distribution function about which we would like to infer; see Section 6 and [3]. Because we require $\mu_U < \infty$ in the above, we restrict our attention to distribution functions G such that $\int z^{-1} \, dG(z) < \infty$.

The connection between the multiplicative censoring model and prevalent cohort studies under the stationarity assumption has revived interest in the former. Nonetheless, there appears to be no result in the literature on density estimation under the multiplicative censoring model, despite its importance in applied sciences. A recent application described by Kvam [28] concerns nanoscience and the measurement of carbon nanotubes. As discussed by Silverman [43], density estimation can be useful for purposes of data exploration and presentation. It is effective in the investigation of modes (determination of multimodality and identification of modes) and tail behavior (rate of tail decay). These features are especially important in length-biased sampling and survival analysis, where skewness is often pervasive and differential subgroup characteristics may lead to multimodality. An additional motivation for the study of density estimation under multiplicative censoring stems from the fact that nonparametric regression of right-censored length-biased data has not been addressed in the literature. In view of the intricate link between density estimation and nonparametric regression (see [35]), a study of density estimation under multiplicative censoring provides foundations for studying nonparametric regression of right-censored length-biased data.

Among the various methods of density estimation, kernel smoothing is particularly appealing for both its simplicity and its interpretability (e.g., as a limiting pointwise average of shifted histograms). It provides a unifying framework in that, as discussed in [40], each of finite difference density estimation, smoothing by convolution, orthogonal series approximations and other smoothing methods historically used in the various applied sciences can be seen as instances of kernel smoothing. This article studies the large-sample properties of kernel density estimators in the setting of multiplicative censoring. Pioneered by Silverman [42], the approach adopted consists of constructing strong approximations of the empirical density process.

Although under the multiplicative censoring model we may avoid complexities altogether by performing estimation using the uncensored observations alone, use of the full data is motivated by at least two reasons. First, although discarding the censored cases under the canonical multiplicative censoring scheme does not compromise consistency, the same cannot be said under the related length-biased sampling scheme, even though these schemes

lead to the same likelihood. This occurs because, under length-bias sampling, the uncensored cases do not emanate directly from the (length-biased version of the) distribution of interest. Systematic exclusion of the censored cases would therefore lead to inconsistency. This fact motivates the study of both censored and uncensored cases under multiplicative censoring. Second, due to the informativeness of the censoring mechanism, ignoring the censored observations may lead to a substantial loss of efficiency. Because the asymptotic covariance function of the nonparametric maximum likelihood estimator of G does not have an explicit form, this phenomenon is difficult to quantify in the nonparametric setting (see the discussion on page 1024 of [51]); however, a parametric example may be illustrative. Suppose that the uncensored observations emanate from a Gamma distribution, say with mean 2θ and variance $2\theta^2$, then the censored observations are exponentially distributed with mean θ. The asymptotic relative efficiency of the full-sample MLE relative to the uncensored-sample MLE is $1+\upsilon/2$, where $\upsilon>0$ is the asymptotic relative frequency of censored observations to uncensored observations. If, for example, $\upsilon=1$, indicating that uncensored and censored cases arise in equal numbers asymptotically, use of the full sample provides a fifty percent gain in efficiency.

Following [27], hereafter referred to as KMT, and [15], we first construct a strong approximation for the process $\sqrt{k}(\hat{G}-G)$, where $\hat{G}$ is a solution of the nonparametric score equation based on $(\mathcal{X}_m,\mathcal{Y}_n)$. The literature on strong approximations is vast. Recent reviews on empirical processes, strong approximations and the KMT construction include [17] and [30]. Using this strong approximation and a result on the global modulus of continuity, we obtain the strong uniform consistency of the kernel density estimators of the density function g associated to G and find a sequence of Gaussian processes strongly uniformly approximating the empirical kernel density process. Using these results, we study the integrated squared error properties of the kernel density estimators.

The layout of the paper is as follows. In Section 2, we introduce our notation and present some preliminaries. In Section 3, we find a sequence of Gaussian processes that strongly uniformly approximates the empirical process $\sqrt{k}(\hat{G}-G)$ and study its global modulus of continuity. We use these results to study the asymptotic behavior of the kernel density estimators in Section 4. It is shown, in particular, that the kernel density estimators are strongly consistent and asymptotically Gaussian. Section 5 is devoted to the integrated squared error properties of the kernel density estimators and includes results from a preliminary small-sample simulation study. We show how our results can be extended to length-biased sampling with right-censoring in Section 6 and present concluding remarks in Section 7. The claim and theorems are proved in the Appendix while lemmas are proved in the supplementary material [1].

2. Preliminaries. We consider the random multiplicative censoring model introduced in [50], whereby two independent random samples $\mathcal{X}_m = (X_1, \ldots, X_m)$ and $\mathcal{Z}_n = (Z_1, \ldots, Z_n)$ are drawn from the lifetime distribution G and a third independent sample $\mathcal{U}_n = (U_1, \ldots, U_n)$, from the standard uniform distribution. Let $Y_i = Z_i U_i$, $i = 1, \ldots, n$, and write $\mathcal{Y}_n = (Y_1, \ldots, Y_n)$. Then $\mathcal{Y}_n$ is a random sample from the absolutely continuous distribution F with density given by (1.1). The observed data consist of $(\mathcal{X}_m, \mathcal{Y}_n)$ while $(\mathcal{Z}_n, \mathcal{U}_n)$ is unobserved.

We begin with the score equation derived from the likelihood $L(G)$ given by (1.2). Let G_m and F_n be, respectively, the empirical distribution functions based on the uncensored observations $x_1, \ldots, x_m$ and the censored cases $y_1, \ldots, y_n$, and write $\hat{p} = m/k$, where $k = m + n$. For simplicity, assume all observations are distinct, and denote by $t_1 < \cdots < t_k$ the values taken by $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$. The distribution function $\hat{G}$ satisfies the nonparametric score equation if, for all $t \geq 0$,

$$d\hat{G}(t) = \hat{p}\, dG_m(t) + (1 - \hat{p}) \left[\int_{0<y\leq t} \frac{dF_n(y)}{\int_{y\leq z} z^{-1}\, d\hat{G}(z)} \right] t^{-1}\, d\hat{G}(t), \tag{2.1}$$

while $\sum_{j=1}^{k} d\hat{G}(t_j) = 1$ and $d\hat{G}(t_j) \geq 0$, $j = 1, \ldots, k$; see [51], page 1025. Integrating both sides of (2.1), we obtain

$$\hat{G}(t) = \hat{p} G_m(t) + (1 - \hat{p}) \int_{0<x\leq t} \left[\int_{0<y\leq x} \frac{dF_n(y)}{\int_{y\leq z} z^{-1} d\hat{G}(z)} \right] x^{-1}\, d\hat{G}(x),$$

where the final integrand is defined to be 0 for $x > t_k$. We say that a sequence of real numbers $\gamma_{m,n}$ satisfies assumption (A0) if

$$\sum_{m,n} G(\gamma_{m,n}) < \infty,$$

where the summation is understood to range over subsample sizes m and n, jointly taken to infinity, so that $\hat{p} \to p \in (0, 1]$. To circumvent problems related to a singularity at the origin, we select a sequence of positive real numbers $\gamma_{m,n}$ satisfying (A0) and consider solutions $\hat{G}$ of (2.1) assigning zero mass below $\gamma_{m,n}$. All results derived in this article apply to any solution of (2.1) with this property. The existence of such solutions is an important fact.

CLAIM 1. *Suppose that* (A0) *holds. Then, for each m and n sufficiently large, (2.1) has a solution $\hat{G}$ such that $\hat{G}(u) = 0$ for each $u < \gamma_{m,n}$.*

If there exists some $\gamma_0 > 0$ such that $G(\gamma_0) = 0$, assumption (A0) is not required. We may simply choose $\gamma_{m,n} = \gamma_0$, and because any solution of (2.1) will have zero mass below $\gamma_{m,n}$, the proposition follows directly from [50].

Define $U_{m,n} = \sqrt{k}(\hat{G} - G)$, $W_{X,m} = \sqrt{m}(G_m - G)$, $W_{Y,n} = \sqrt{n}(F_n - F)$, $\hat{f}(t) = \int_{t \le z} z^{-1}\, d\hat{G}(z)$ and

$$W_{m,n}(t) = \sqrt{\hat{p}} W_{X,m}(t) + \sqrt{1-\hat{p}}\hat{f}(t) \int_{0<y\le t} W_{Y,n}(y)\, d\left[\frac{1}{\hat{f}(y)}\right]. \tag{2.2}$$

We observe, in particular, that

$$|W_{m,n}(t)| \le \sqrt{\hat{p}}|W_{X,m}(t)| + \sqrt{1-\hat{p}} \sup_{0<y\le t} |W_{Y,n}(y)| \tag{2.3}$$

for each $t > 0$. As in [51], we have that

$$W_{m,n}(t) = \hat{p} U_{m,n}(t) + (1-\hat{p})\hat{f}(t) \int_{0<y\le t} y\left(\int_{y\le z} \frac{U_{m,n}(z)}{z^2}\, dz\right) d\left[\frac{1}{\hat{f}(y)}\right].$$

The process $W_{m,n}$ can therefore be expressed as the image of a linear operator applied on $U_{m,n}$. To see this, we define the operator $\mathcal{G}_{m,n}$ pointwise as $\mathcal{G}_{m,n}(u)(t) = \hat{f}(t)\mathcal{A}_{m,n}(u)(t)$, where

$$\mathcal{A}_{m,n}(u)(t) = \int_{0<y\le t} y\left(\int_{y\le z} \frac{u(z)}{z^2}\, dz\right) d\left[\frac{1}{\hat{f}(y)}\right].$$

Then, we may write $\mathcal{F}_{m,n} = \hat{p}\mathcal{I} + (1-\hat{p})\mathcal{G}_{m,n}$, with $\mathcal{I}(u) = u$ the identity map, and observe that

$$W_{m,n} = \mathcal{F}_{m,n}(U_{m,n}). \tag{2.4}$$

Denoting by $D_0[0,\infty]$ the space of *cadlag* functions vanishing at 0 and ∞ endowed with the uniform topology (the topology induced by the supremum norm over $[0,\infty)$, $\|u\|_\infty = \sup_{0\le t<\infty}|u(t)|$), it is not difficult to see that $\mathcal{I}$, $\mathcal{G}_{m,n}$ and $\mathcal{F}_{m,n}$ are bounded linear operators on $D_0[0,\infty]$, and, in view of Lemma 3 of [51], that $\mathcal{F}_{m,n}$ has a bounded inverse satisfying $\|\mathcal{F}_{m,n}^{-1}\| \le 2/\hat{p}^2$. As in [51], it holds that if $\hat{p} \to p \in (0,1]$ as $m, n \to \infty$, then, for each $u \in D_0[0,\infty]$, we have that

$$\|\mathcal{F}_{m,n}(u) - \mathcal{F}(u)\|_\infty \xrightarrow{\text{a.s.}} 0,$$

where the limit operators are $\mathcal{F} = p\mathcal{I} + (1-p)\mathcal{G}$, $\mathcal{G}(u)(t) = f(t)\mathcal{A}(u)(t)$ and

$$\mathcal{A}(u)(t) = \int_{0<y\le t} y\left(\int_{y\le z} \frac{u(z)}{z^2}\, dz\right) d\left[\frac{1}{f(y)}\right].$$

We may then conclude that $\mathcal{G}$ and $\mathcal{F}$ are also bounded linear operators on $D_0[0,\infty]$ and that $\mathcal{F}$ has a bounded inverse satisfying $\|\mathcal{F}^{-1}\| \le 2/p^2$. Vardi [51] proved the uniform strong consistency of $\hat{G}$ using (2.4). Instead, we obtain it as a corollary of Lemma 1 below.

Of importance will be the fact, proved in [51], that the inverse operator $\mathcal{F}^{-1}$ has the following pointwise representation:

$$\mathcal{F}^{-1}(u)(t) = p^{-1}u(t) + \int_0^\infty \Phi(t,x)u(x)\,dx \tag{2.5}$$

with kernel Φ satisfying, for each t and x, the constraints

$$p^2\Phi(t,x) + (1-p)\mathcal{A}_0(t,x) + p(1-p)\int_0^\infty \Phi(t,z)\mathcal{A}_0(z,x)\,dz = 0 \tag{2.6}$$

and

$$\int_0^\infty \Phi(t,z)\mathcal{A}_0(z,x)\,dz = \int_0^\infty \mathcal{A}_0(t,z)\Phi(z,x)\,dz, \tag{2.7}$$

where we have defined $\mathcal{A}_0(t,x) = f(t)x^{-2}\int_{0<y\le t\wedge x} y\,d[1/f(y)]$.

As in [51], we have that $W_{m,n} \rightsquigarrow W$ in $D_0[0,\infty]$, where W is the Gaussian process

$$W(t) = \sqrt{p}B_X(G(t)) + \sqrt{1-p}f(t)\int_{0<y\le t} B_Y(F(y))\,d\left[\frac{1}{f(y)}\right]$$

with B_X and B_Y independent Brownian bridges, and that $U_{m,n} \rightsquigarrow U = \mathcal{F}^{-1}(W)$ in $D_0[0,\infty]$. Here, the symbol $\rightsquigarrow$ refers to weak convergence. This last step can be established using the convergence of $\mathcal{F}_{m,n}$ to $\mathcal{F}$ in operator norm topology, Lemma 3 of [51] and the continuous mapping theorem. A consistent estimator $\hat{\psi}_U(s,t)$ of $\psi_U(s,t) = \mathrm{E}[U(s)U(t)]$ is provided in [51], though in practice the use of resampling methods may yield an estimator of $\psi(s,t)$ more expediently.

3. Approximation of the empirical process $U_{m,n}$.

3.1. *Strong approximation.* Let α_n denote the empirical process of n independent standard uniform random variables. The KMT construction implies that there exists a probability space $(\Omega,\mathcal{F},P)$ with a sequence of independent standard uniform random variables and a sequence of Brownian bridges B_n such that

$$\|\alpha_n - B_n\|_{[0,1]} = \mathcal{O}\left(\frac{\log n}{\sqrt{n}}\right) \qquad \text{a.s.}$$

Equation (2.4) is key to the strong approximation of $U_{m,n}$. Since $W_{X,m}$ and $W_{Y,n}$ are independent empirical processes associated, respectively, with $\mathcal{X}_m$ and $\mathcal{Y}_n$, in view of the KMT construction, there exist versions of $W_{X,m}$ and $W_{Y,n}$ along with two independent sequences of Brownian bridge processes $B_{X,m}$ and $B_{Y,n}$ such that $B_{X,m}\circ G$ and $B_{Y,n}\circ F$ approximate $W_{X,m}$

and $W_{Y,n}$ at the optimal rate of $\log s/\sqrt{s}$ (here, s is the sample size). Using (2.4), we extend this approximation to $W_{m,n}$ and use properties of $\mathcal{F}$ to find a sequence of Gaussian processes strongly uniformly approximating $U_{m,n}$. The main theorem of this section, Theorem 1, is proved through a sequence of lemmas.

Denote the upper limit of the support of G by $\tau = \sup\{t : G(t) < 1\}$. Given any set B, denote by $\mathbb{I}_B$ and $\|\cdot\|_B$ the indicator function and the supremum norm over B, respectively. Write $\|\cdot\|_\infty$ for the case $B = [0, \infty)$. We introduce the following assumptions:

(A1) $\sqrt{k}(\hat{p} - p) = \mathcal{O}(\sqrt{\log\log k})$ for some $p \in (0, 1]$.
(A2) G is continuous and has bounded support ($\tau < \infty$).
(A3) There exists $\alpha_0 > 2$ such that $\lim_{x\downarrow 0} G(x)/x^{\alpha_0} < \infty$.
(A4) There exists $\beta > 0$ such that $\lim_{x\downarrow 0}[1 - G(\tau - x)]/x^\beta \in (0, \infty)$.

We begin by obtaining rates for the difference between $\hat{G}$ and G as well as between $\hat{f}$ and f in the supremum norm.

LEMMA 1. *Suppose* (A0) *holds. Then, for any sequence of nonnegative real numbers $a_{m,n}$, as $k \to \infty$:*

$$\text{(a)} \quad \|\hat{G} - G\|_\infty = \mathcal{O}\left(\sqrt{\frac{\log\log k}{k}}\right) \qquad a.s.,$$

$$\text{(b)} \quad \|\hat{f} - f\|_{[a_{m,n},\infty)} = \mathcal{O}\left(\gamma_{m,n}^{-1}\sqrt{\frac{\log\log k}{k}} + [F_U(\gamma_{m,n}) - F_U(a_{m,n})]\mathbb{I}_{[0,\gamma_{m,n})}(a_{m,n})\right) \quad a.s.$$

The above indicates, for example, that in addition to satisfying (A0), $\gamma_{m,n}$ should be such that

$$\gamma_{m,n}^{-1}\sqrt{\frac{\log\log k}{k}} \to 0.$$

If (A3) holds, the sequence $\gamma'_{m,n} = k^{-1/(2\alpha)}$ may be considered, with the choice $\alpha \in (1, \alpha_0/2)$ ensuring that the two requirements above are satisfied. In this case, choosing α as close as possible to $\alpha_0/2$ would yield the fastest rate, modulo logarithmic terms, in part (b) of Lemma 1. We now provide a result on the growth rate of maxima of Wiener processes.

LEMMA 2. *Let $\mathcal{W}_n$ be a sequence of standard Wiener processes. Then, as $n \to \infty$,*

$$\|\mathcal{W}_n\|_{[0,1]} = \mathcal{O}(\sqrt{\log n}) \qquad a.s.$$

The next result considers the asymptotic behavior of the sequence of inverse operators $\mathcal{F}_{m,n}^{-1}$. First, we note that the space $D_0[0,\tau]$ endowed with the uniform topology is a Banach space. As such, $\mathscr{A} = \mathcal{L}(D_0[0,\tau], D_0[0,\tau])$, the space of bounded linear operators on $D_0[0,\tau]$ endowed with the operator norm topology, is a Banach algebra. We recall additionally that *cadlag* functions have countably many jumps (see [36]) and are therefore Riemann integrable on bounded intervals.

Fixing $\varepsilon > 0$, set $\mathcal{I}_\varepsilon(u)(t) = u(t)\mathbb{I}_{[0,\tau-\varepsilon]}(t)$ and define $\mathcal{F}_{m,n,\varepsilon}$ and $\mathcal{F}_\varepsilon : D_0[0,\tau] \to D_0[0,\tau]$ as

$$\mathcal{F}_{m,n,\varepsilon} = \hat{p}\mathcal{I} + (1-\hat{p})\mathcal{G}_{m,n,\varepsilon} \quad \text{and} \quad \mathcal{F}_\varepsilon = p\mathcal{I} + (1-p)\mathcal{G}_\varepsilon,$$

respectively, where for any $t \in [0,\tau]$,

$$\mathcal{G}_{m,n,\varepsilon}(u)(t) = \hat{f}(t)(\mathcal{A}_{m,n} \circ \mathcal{I}_\varepsilon)(u)(t) \quad \text{and} \quad \mathcal{G}_\varepsilon(u)(t) = f(t)(\mathcal{A} \circ \mathcal{I}_\varepsilon)(u)(t).$$

Define $\varepsilon_0 = \tau p^2/(p^2 - 2p + 2)$.

Lemma 3. *Suppose that* (A0)–(A2) *hold and that ε is in $(0,\varepsilon_0)$. Then, considering the operator norm over the space $C_0[0,\tau]$ of continuous functions on $[0,\tau]$ vanishing at the endpoints, as $k \to \infty$,*

$$\begin{aligned}\|\mathcal{F}_{m,n,\varepsilon}^{-1} - \mathcal{F}_\varepsilon^{-1}\| &\\ = \mathcal{O}&\left(\left[\frac{\log(1/\gamma_{m,n})}{f(\tau-\varepsilon)} + \frac{F_U(\gamma_{m,n})}{f(\gamma_{m,n})}\right]\gamma_{m,n}^{-1}\sqrt{\frac{\log\log k}{k}} + F_U(\gamma_{m,n})\right) \qquad a.s.\end{aligned}$$

With the choice $\gamma_{m,n} = \gamma'_{m,n}$, the order above may be simplified to

$$\|\mathcal{F}_{m,n,\varepsilon}^{-1} - \mathcal{F}_\varepsilon^{-1}\| = \mathcal{O}\left(\frac{k^{-(\alpha-1)/(2\alpha)}\log k\sqrt{\log\log k}}{f(\tau-\varepsilon)}\right) \qquad \text{a.s.}$$

We now consider a random integral useful in determining the rate of the strong approximation we will construct for $U_{m,n}$.

Lemma 4. *Suppose that* (A0)–(A2) *hold and that ε is in $(0,\varepsilon_0)$. Then, as $k \to \infty$,*

$$\begin{aligned}\sup_{0\le s\le\tau-\varepsilon} \hat{f}(s)\left|\int_0^s B_{Y,n}(F(y))\, d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right]\right| &\\ = \mathcal{O}\left(\frac{k^{-1/4}\sqrt{\log k}(\log\log k)^{1/4}}{f(\tau-\varepsilon)}\right) \qquad a.s.\end{aligned}$$

Remark 1. The above bound also holds for $\varepsilon = \varepsilon_{m,n} \downarrow 0$ provided $\varepsilon_{m,n}k/\sqrt{\log\log k} \to \infty$.

Henceforth, we set $\gamma_{m,n}=\gamma'_{m,n}$ for each m and n. The next lemma establishes the existence of a sequence of Gaussian processes approximating $W_{m,n}$. Define the sequence of processes

$$W_{m,n}^0(s)=\sqrt{p}B_{X,m}(G(s))+\sqrt{1-p}f(s)\int_{0<y\le s}B_{Y,n}(F(y))\,d\left[\frac{1}{f(y)}\right]. \tag{3.1}$$

LEMMA 5. *Suppose that* (A1)–(A3) *hold and that* ε *is in* $(0,\varepsilon_0)$. *Then, setting there exists a probability space on which* $W_{m,n}$ *and* $W_{m,n}^0$ *are defined such that, as* $k\to\infty$,

$$\|W_{m,n}-W_{m,n}^0\|_{[0,\tau-\varepsilon]}=\mathcal{O}\left(\frac{k^{-r(\alpha)}\sqrt{\log k}(\log\log k)^{1/4}}{f(\tau-\varepsilon)}\right)\qquad a.s.,$$

where $r(\alpha)=\min(\frac{1}{4},\frac{\alpha-1}{2\alpha})$.

The next lemma extends the result on the growth rate of Wiener processes in Lemma 2 to the sequence of approximating processes (3.1).

LEMMA 6. *Suppose that* (A2) *holds and that* $p\in(0,1]$. *Then, as* $k\to\infty$,

$$\|W_{m,n}^0\|_\infty=\mathcal{O}(\sqrt{\log k})\qquad a.s.$$

Having established the existence of a sequence $W_{m,n}^0$ of Gaussian processes approximating $W_{m,n}$ and studied the behavior of $\mathcal{F}_{m,n}^{-1}$, we may provide a sequence of Gaussian processes approximating $U_{m,n}$. Define $U_{m,n}^0=\mathcal{F}^{-1}(W_{m,n}^0)$ for each m and n. Since $\mathcal{F}^{-1}$ is a bounded linear operator, $U_{m,n}^0$ forms a sequence of Gaussian processes.

THEOREM 1. *Suppose that* (A1)–(A4) *hold. Then, on the probability space on which* $W_{m,n}$ *and* $W_{m,n}^0$ *are defined, we have that, as* $k\to\infty$,

$$\|U_{m,n}-U_{m,n}^0\|_{[0,\tau-\varepsilon_{m,n}]}=\mathcal{O}(\varepsilon_{m,n}(\log k)^{3/2}\sqrt{\log\log k})\qquad a.s.,$$

where $\varepsilon_{m,n}=k^{-r(\alpha)/(\beta+1)}$ *and* $r(\alpha)=\min(\frac{1}{4},\frac{\alpha-1}{2\alpha})$.

Theorem 1 will be crucial in our study of the asymptotic properties of kernel density estimators of g, the density associated to G, in Sections 4 and 5. Other applications of Theorem 1 include oscillation moduli and laws of the iterated logarithm; see [16].

3.2. *Global modulus of continuity.* In order to describe the asymptotic properties of the kernel density estimators of g via the above strong approximation, we must establish the global modulus of continuity of the approximating process $U_{m,n}^0$.

In the sequel, we say that the distribution G satisfies assumption (A5) if its density g is differentiable, and that a sequence of bandwidths $h_{m,n}$

satisfies assumption (B1) if:

(1) $mh_{m,n} \to \infty$ and $\log h_{m,n}/\log\log m \to -\infty$ as $m,n \to \infty$;
(2) $\sqrt{\log n} h_{m,n} \to 0$ and $\sqrt{\log m} h_{m,n} \to 0$ as $m,n \to \infty$.

THEOREM 2. *Suppose that* (A1)–(A5) *hold, and that the sequence* $h_{m,n}$ *satisfies* (B1). *Then, for any* η *in* $(0,\tau)$, *we have that, as* $k \to \infty$,

$$\sup_{0 \le t \le \tau-\eta} \sup_{0 \le s \le h_{m,n}} |U^0_{m,n}(t+s) - U^0_{m,n}(t)| = \mathcal{O}(\sqrt{h_{m,n}\log(1/h_{m,n})}) \qquad a.s.$$

4. Asymptotic behavior of kernel density estimators. Consider the kernel density estimator $\hat{g}_m$ of a univariate density g introduced by [38],

$$\hat{g}_m(t) = \frac{1}{h_m}\int_0^\infty K\left(\frac{t-s}{h_m}\right) d\hat{G}_m(s), \tag{4.1}$$

where $X_1,\ldots,X_m$ are independent observations from g, K is some kernel function, h_m some bandwidth and $\hat{G}_m$ the empirical distribution function. The weak and strong uniform consistency of $\hat{g}_m$ was addressed in [33, 39] and [47], among others. To ensure strong uniform consistency, these authors required that $\sum_m \exp(-cm{h_m}^2) < \infty$ for each $c > 0$. Silverman [42] established the strong uniform consistency of $\hat{g}_m$ under weaker assumptions using the KMT strong approximation technique. When the observations are subject to random right-censoring, Blum and Susarla [9] proposed estimating g by the estimator in (4.1), replacing $\hat{G}_m$ by the Kaplan–Meier estimator of G. The properties of the resulting estimator were examined in [9, 19] and [32], among others.

To estimate the density function g under multiplicative censoring, we consider a sequence of kernel density estimators $\hat{g}_{m,n}$, defined as

$$\hat{g}_{m,n}(t) = \frac{1}{h_{m,n}}\int_0^\infty K\left(\frac{t-s}{h_{m,n}}\right) d\hat{G}(s), \tag{4.2}$$

where $\hat{G}$ is, as before, a solution of the nonparametric score equation based on $(\mathcal{X}_m, \mathcal{Y}_n)$.

We introduce an additional set of assumptions to be used in the sequel. The sequence of bandwidths $h_{m,n}$ is said to satisfy assumption (B2) if

$$\lim_{k\to\infty} \frac{\varepsilon_{m,n}(\log k)^{3/2}\sqrt{\log\log k}}{\sqrt{k} h_{m,n}} = 0.$$

We say that a kernel function K satisfies assumption (K1) if:

(1) K has total variation $V_K < \infty$;
(2) K is supported on $(-1,1)$;
(3) K is continuous;
(4) $\int K(u)\,du = 1$.

Further, we say that it satisfies assumption (K2) if $\int uK(u)\,du = 0$.

4.1. *Strong uniform consistency.* Denote by $g_{m,n}$ the kernel smoothing of g based on G; that is, write

$$g_{m,n}(t) = \frac{1}{h_{m,n}} \int_0^\infty K\left(\frac{t-s}{h_{m,n}}\right) dG(s).$$

LEMMA 7. *Suppose that* (A1)–(A5) *hold, and that $h_{m,n}$ is a sequence of positive bandwidths tending to 0 as $k \to \infty$ and satisfying* (B1) *and* (B2). *Suppose also that the kernel function K satisfies* (K1). *Then, for any η in $(0, \tau)$, we have that*

$$\lim_{k\to\infty} \|\hat{g}_{m,n} - g_{m,n}\|_{[0,\tau-\eta]} = 0 \qquad a.s.$$

THEOREM 3. *Suppose that* (A1)–(A5) *hold, and that $h_{m,n}$ is a sequence of positive bandwidths tending to 0 as $k \to \infty$ and satisfying* (B1) *and* (B2). *Suppose also that the kernel function K satisfies* (K1). *Then, for any η in $(0, \tau)$, we have that*

$$\lim_{k\to\infty} \|\hat{g}_{m,n} - g\|_{[0,\tau-\eta]} = 0 \qquad a.s.$$

4.2. *Strong uniform approximation of the empirical density process.* By Theorems 1 and 3, we can find a sequence of Gaussian processes that strongly and uniformly approximates the empirical density process. Let K be an arbitrary density function, and define

$$\varphi_{m,n}(t,s) = \frac{1}{h_{m,n}} K\left(\frac{t-s}{h_{m,n}}\right).$$

Denoting by $\mathrm{v}_s[\varphi_{m,n}(t,s)]$ the total variation of $\varphi_{m,n}(t,\cdot)$ for fixed t, we refer to the uniform total variation $\sup_t \mathrm{v}_s[\varphi_{m,n}(t,s)]$ by $V_{m,n}$.

THEOREM 4. *Suppose that* (A1)–(A5) *hold, and that $h_{m,n}$ is a sequence of positive bandwidths tending to 0 as $k \to \infty$ and satisfying* (B1) *and* (B2). *Suppose also that the kernel function K satisfies* (K1) *and* (K2), *and that g has a bounded second derivative. Then, for any η in $(0, \tau)$, we have that*

$$\begin{aligned} &\|\sqrt{k}(\hat{g}_{m,n} - g) - \Gamma_{m,n}\|_{[0,\tau-\eta]} \\ &\quad = \mathcal{O}\left(\frac{\varepsilon_{m,n}(\log k)^{3/2}\sqrt{\log\log k}}{h_{m,n}} + \sqrt{k}h_{m,n}^2\right) \qquad a.s., \end{aligned}$$

where we have defined $\Gamma_{m,n}(t) = \int_0^\infty U_{m,n}^0(s)\varphi_{m,n}(t,ds)$.

REMARK 2. Theorem 4 suggests that the optimal rate for the above approximation is obtained by choosing $h_{m,n} \sim (\varepsilon_{m,n}\sqrt{\log\log k/k})^{1/3}\sqrt{\log k}$.

Theorem 4 implies distributional results. The linearization $\psi_U(s-uh, t-vh) - \psi_U(s,t) \sim h$ is useful here. This result is not difficult to show for $p > 1/2$

using representations of ψ_U provided on page 1033 of [51], linearization techniques and the modulus of continuity of process U. The case $p \le 1/2$ (i.e., heavy censoring) is more challenging, but can be dealt with using (2.6), (2.7) and an argument similar to that found in the proof of Theorem 2. Using Theorem 4 and the above linearization, we may show that $\sqrt{kh_{m,n}}(\hat{g}_{m,n} - g)$ is asymptotically Gaussian with mean zero and covariance function σ_g estimated consistently by

$$\hat{\sigma}_g(s,t) = h_{m,n}^{-1} \int\int \hat{\psi}_U(s - uh_{m,n}, t - vh_{m,n})\, dK(u)\, dK(v).$$

5. Integrated squared error of kernel density estimators. A common measure of the global performance of an estimator $\hat{g}_m$ of a density g is its integrated square error (ISE), defined as

$$\mathcal{E}_m = \int_{-\infty}^{\infty} [\hat{g}_m(s) - g(s)]^2\, ds.$$

Use of the ISE is particularly pervasive in simulation studies aiming to compare the performance of various density estimators. Minimization of the mean integrated square error (MISE) $\mathrm{E}[\mathcal{E}_m] = \int_{-\infty}^{\infty} \mathrm{E}[g_m(s) - g(s)]^2\, ds$ is often a guiding principle in the construction of kernel density estimators. Steele [44] identified the need to determine the relationship between various measures of accuracy in density estimation. One such measure, the order of $\mathcal{E}_m - \mathrm{E}(\mathcal{E}_m)$, is particularly important in statistics. Hall [24] first began addressing the issues raised in [44] by computing the exact order of convergence of $\mathcal{E}_m - \mathrm{E}(\mathcal{E}_m)$ to zero using the strong approximation technique developed by Komlós, Major and Tusnády [27] for the standard empirical process. Zhang [56] studied the case of random right-censoring using the strong approximation technique of [10] and [11]. In this section, we consider the ISE $\mathcal{E}_{m,n}$ of the kernel estimator $\hat{g}_{m,n}$ under multiplicative censoring and derive its asymptotic expansion.

5.1. *Asymptotic expansion of the integrated squared error.* In the remainder of the paper, we make use of the following assumptions. We say that the kernel function K satisfies assumption (K3) if it has finite second moment $\sigma^2 > 0$ and is differentiable. Further, we say that the density g satisfies assumption (A6) if it is twice continuously differentiable. Of course, assumption (A6) implies assumption (A5). Finally, we say that the sequence of bandwidths $h_{m,n}$ satisfies assumption (B3) if

$$\lim_{k \to \infty} \frac{\sqrt{\log k}(\log\log k)^{1/6}}{h_{m,n} k^{1/(\delta(\beta))}} = 0,$$

where $\delta(\beta) = 4 + 4\beta/(2\beta + 3)$. In the sequel, we write ν for $\sqrt{\int_{-1}^{1} K^2(u)\, du}$.

The ISE of $\hat{g}_{m,n}$ on the interval $[u_1, u_2]$ is defined as

$$\mathcal{E}_{m,n}(u_1, u_2) = \int_{u_1}^{u_2} [\hat{g}_{m,n}(s) - g(s)]^2 \, ds.$$

Theorem 5 presents an asymptotic expansion for $\mathcal{E}_{m,n}(0, \tau - \eta)$ for any η in $(0, \tau)$.

THEOREM 5. *Suppose that* (A1)–(A4) *and* (A6) *hold with* $\alpha_0 > 4$ *in* (A3) *and that* α *is chosen in* $[2, \alpha_0/2)$. *Suppose that* $h_{m,n}$ *is a sequence of positive bandwidths satisfying* (B1) *and* (B3), *and that the kernel function* K *satisfies* (K1)–(K3). *Then, for any* η *in* $(0, \tau)$, *we have that*

$$\mathcal{E}_{m,n}(0, \tau - \eta) = \frac{h_{m,n}^4 \sigma^4}{4} \int_0^{\tau-\eta} [g''(s)]^2 \, ds + \frac{\nu^2}{h_{m,n} kp} + o_p\left(\frac{1}{kh_{m,n}} + h_{m,n}^4\right).$$

Theorem 5 suggests that $h_{m,n}$ should shrink at the rate $k^{-1/(\zeta(\beta))}$ modulo logarithmic terms, where $\zeta(\beta) = \max(5, \delta(\beta))$. We note that $\delta(\beta) < 5$ when $\beta < 3/2$. Then, writing $\|g''\|_{2,[0,\tau-\eta]}^2 = \int_0^{\tau-\eta} [g''(s)]^2 \, ds$, Theorem 5 suggests that the bandwidth

$$h_{m,n}^{\star} = \left(\frac{\nu^2}{kp\sigma^4 \|g''\|_{2,[0,\tau-\eta]}^2}\right)^{1/5}$$

minimizes the order of the integrated squared error, a direct generalization of the reference rule for uncensored data alone, which we recover for $p = 1$ and $k = m$. Of course, in practice, this bandwidth is unknown; instead, we may substitute g'' by some estimate $\hat{g}''$, and p by $\hat{p} = m/k$. For example, a reference rule based on a Gamma approximation to G is given by

$$\hat{h}_{m,n}^{\star} = 2\hat{\beta}\left(\frac{\nu^2}{m\sigma^4}\right)^{1/5}, \tag{5.1}$$

where $\hat{\beta} = \sum_{i=1}^m X_i/(4m)$ is the MLE of β based on $\mathcal{X}_m$ and the model $G = G_\beta$, with $g_\beta(x) = x^3 \exp(-x/\beta)/(6\beta^4)$ the density associated to G_β. This distribution satisfies (A3) with $\alpha_0 = 4$ but is a limiting case with respect to the stronger assumption made in Theorem 5. It was selected because it has the least smooth density in the family of densities $\{g_{\alpha,\beta}(x) = x^{\alpha-1} \exp(-x/\beta)/[\Gamma(\alpha)\beta^\alpha] : \alpha \geq 4\}$ with respect to the L_2-norm of the second derivative of $g_{\alpha,\beta}$. Alternatively, we may consider kernel smoothing of the uncensored observations alone to obtain a nonparametric pilot estimate $\hat{g}''$ of g''. More robust but computationally intensive cross-validation approaches, as in [29], may also be used for bandwidth selection.

5.2. *Small-sample simulation results: Implementation and efficiency.* To provide some illustration of the behavior of the methods proposed, we present below results from a preliminary small-sample simulation study. The objec-

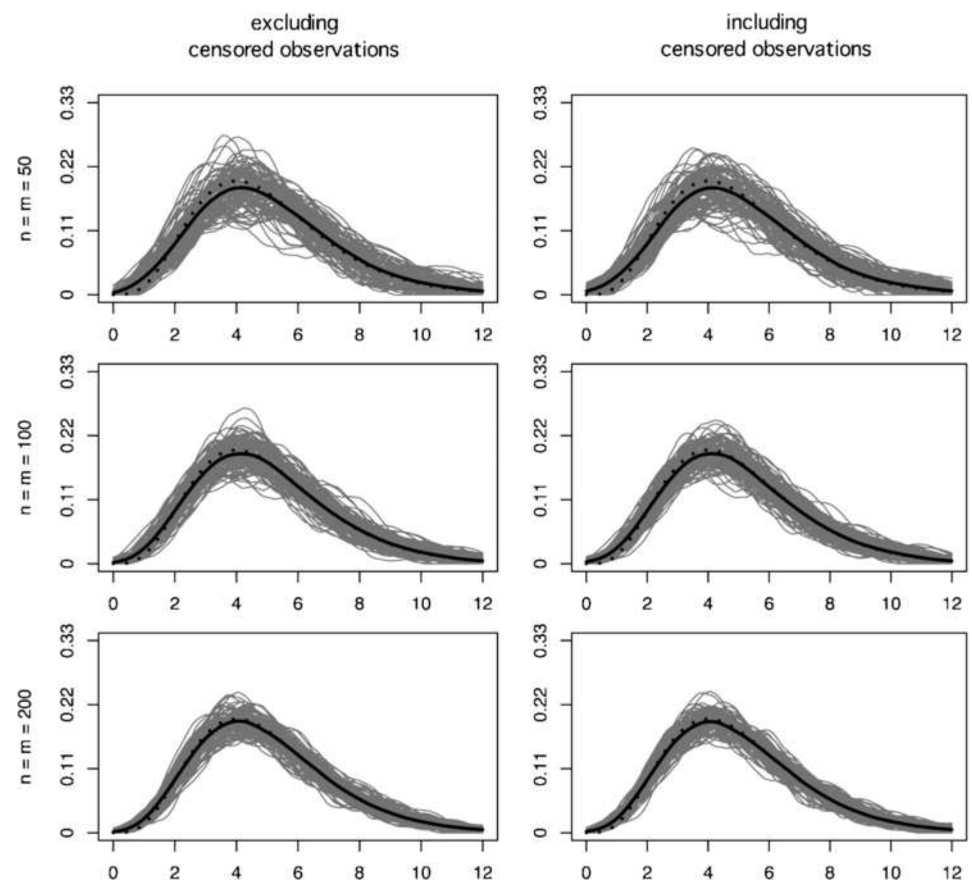

FIG. 1. *Overlayed sample paths.*

tive was to graphically evaluate the general adequacy of the estimators as well as to elucidate the potential contribution of censored observations to overall estimation efficiency, both in small samples. For this purpose, we considered data emanating from the multiplicative censoring model, with underlying Gamma density function $g_\alpha(x) = \Gamma(\alpha)^{-1}x^{\alpha-1}\exp(-x)\mathbb{I}_{(0,\infty)}(x)$, various sample sizes and differing values of parameter α. We found the kernel density estimators proposed to perform generally well. Figure 1 presents 100 sample paths, shown in grey, for various sample sizes and parameter value $\alpha = 5$. Plots in the first column were obtained by discarding all censored observations and performing kernel density estimation using the uncensored observations alone; all observations were used in generating plots in the second column. The pointwise average of the sample plots is shown in solid black, while the true density is the dotted black curve depicted. The first, second and third rows were generated from datasets of 100, 200 and 400

Table 1
Average percent increase in ISE and 95% CIs using $\Gamma(4,\beta)$ *parametric reference rule*

Sample size	$\boldsymbol{\alpha=3}$	$\boldsymbol{\alpha=4}$	$\boldsymbol{\alpha=5}$	$\boldsymbol{\alpha=6}$
$50+50$	${}_{15.2}17.1_{19.0}$	${}_{11.3}13.4_{15.6}$	${}_{16.4}18.4_{20.4}$	${}_{13.9}16.3_{18.7}$
$100+100$	${}_{16.8}18.5_{20.2}$	${}_{14.1}15.8_{17.5}$	${}_{13.6}15.3_{17.0}$	${}_{9.9}11.6_{13.2}$
$200+200$	${}_{13.2}14.7_{16.1}$	${}_{11.6}13.1_{14.7}$	${}_{14.4}17.8_{21.2}$	${}_{18.4}22.6_{26.7}$

total observations, respectively, with censored and uncensored observations equally represented. In all cases, bandwidth values were automatically selected using the $\Gamma(4,\beta)$ parametric reference rule (5.1). The Epanechnikov kernel $K(x)=\frac{3}{4}(1-x^2)\mathbb{I}_{(-1,1)}(x)$ was used throughout. From these plots, we notice that use of the full sample leads to a decrease in variability throughout the support. Our empirical findings suggest that this cumulates to a substantial decrease in integrated squared error. Table 1 reports estimates and associated 95% confidence intervals for the mean relative difference in ISE, defined as $(\mathrm{ISE}_0-\mathrm{ISE}_1)/\mathrm{ISE}_1$, obtained from a simulation of 500 datasets, where ISE_0 and ISE_1 are the integrated squared errors associated with the use of the uncensored subsample and of the full sample, respectively. These values describe the mean percent increase in ISE from discarding the censored subsample, for various sample sizes and parameter values.

The relative performance of the estimators was found to be rather insensitive to the proximity of the underlying distribution to the parametric model specified in the reference rule used, with an average increase in ISE of around 10–25%, subsequent to discarding censored observations, regardless of sample size and parameter value. Since the performance of kernel density estimators hinges upon the performance of the underlying estimator of the distribution function as well as the adequacy of the bandwidth selection rule, gauging the contribution of censored observations to overall estimation efficiency is complicated by the layer of uncertainty associated to bandwidth selection. As such, we have also conducted a simulation study, whereby, for each simulated dataset, the bandwidth selected was that minimizing the observed ISE; we refer to this rule as the optimal bandwidth selection rule. Of course, such a rule can only be adopted in simulation settings, where the true density function is known, and the ISE can be computed directly. This approach provides, nonetheless, a clearer view of the gains resulting from the inclusion of censored observations in the estimation procedure. Table 2 reports estimates of the mean relative increase in ISE resulting from discarding all censored observations along with associated 95% confidence intervals. These results seem to suggest that for small and moderate sample sizes, when equal numbers of censored and uncensored observations are available, ignoring censored observations leads to an increase in ISE of roughly 10–35%, results consistent with those reported in Table 1.

TABLE 2
Average percent increase in ISE and 95% CIs using optimal bandwidth selection rule

Sample size	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$	$\alpha = 6$
$50 + 50$	${}_{9.6}14.3_{19.0}$	${}_{10.9}15.7_{20.5}$	${}_{9.8}14.8_{19.9}$	${}_{17.3}32.5_{47.7}$
$100 + 100$	${}_{12.7}16.3_{20.0}$	${}_{12.9}17.3_{21.6}$	${}_{11.1}15.8_{20.5}$	${}_{12.4}26.9_{41.3}$
$200 + 200$	${}_{10.0}13.8_{17.6}$	${}_{12.1}17.2_{22.4}$	${}_{14.3}21.0_{27.7}$	${}_{16.9}34.6_{52.3}$

The above provides a glimpse of the contribution of the censored observations in small and moderate samples. It suggests that these observations provide nonnegligible information regarding the estimand of interest. We may, however, also resort to asymptotic arguments to motivate use of the full sample for the sake of efficiency. For any given distribution function H, denote the integrated squared error by

$$\mathrm{ISE}(H,h;g) = \int \left[\frac{1}{h} \int K\left(\frac{x-y}{h}\right) dH(y) - g(y) \right]^2 dy$$

and define the optimal bandwidth $\lambda(H;g)$ as the minimizer of the ISE with respect to the true density g, that is, $\lambda(H;g) = \arg\min_{h>0} \mathrm{ISE}(H,h;g)$. Let $G_{m,n}$ be any consistent estimator of G based on $(\mathcal{X}_m, \mathcal{Y}_n)$. The optimal kernel density estimator of g based on $G_{m,n}$ is then $g^\star_{m,n} = \omega(G_{m,n})$, where ω is the operator defined pointwise as

$$\omega(H)(x) = \frac{1}{\lambda(H;g)} \int K\left(\frac{x-u}{\lambda(H;g)}\right) dH(u).$$

Since any solution $\hat{G}$ of the nonparametric score equation is asymptotically efficient for G (see [51]), it is possible to show, along the lines of Theorem 25.47 of [46], that $\hat{g}^\star_{m,n} = \omega(\hat{G})$ is asymptotically efficient for $g = \omega(G)$. In particular, the kernel density estimator using the empirical distribution function based on uncensored observations alone cannot be expected to be asymptotically efficient, given that the latter is itself not efficient for G. It is thus clear that, barring additional complications linked to bandwidth selection, use of the full sample is preferable to that of the uncensored subsample alone.

6. Length-biased sampling with right-censoring. As discussed in the Introduction, the likelihood of length-biased right-censored data is a particular case of that exhibited in (1.2). The literature on length-biased sampling can be traced as far back as [52], with important contributions by Fisher [18], Neyman [34] and Zelen [55] in medical applications, and by Cox [13] in industrial applications. The rigorous treatment of biased sampling was initiated in the 1980s by Vardi [48, 49], and furthered by Gill, Vardi and Wellner [21], Vardi and Zhang [51], Bickel and Ritov [7], Gilbert [20]

and, more recently, by Asgharian, M'Lan and Wolfson [2], Asgharian and Wolfson [3] and Bergeron, Asgharian and Wolfson [5]. The importance of biased sampling in medical applications and prevalent cohort studies was re-emphasized by Cox and Oakes [14].

The lifetime data typically collected on a prevalent cohort consist of triples $(A, R\wedge D, \Delta)$, where A, R and D are, respectively, the current-age, the residual lifetime and the residual censoring time, while $\Delta = \mathbb{I}_{\{R\leq D\}}$ is the censoring indicator. Suppose that D and (A,R) are independent. In one scenario considered in [3], all analyses are carried out conditionally upon the proportion of uncensored individuals, assumed fixed. As such, the observations are comprised of

$$(A_i, R_i) \overset{\text{i.i.d.}}{\sim} f_{A,R|\Delta=1}, \qquad i = 1, \dots, m,$$

and

$$(A_j, D_j) \overset{\text{i.i.d.}}{\sim} f_{A,D|\Delta=0}, \qquad j = 1, \dots, n,$$

where $f_{A,R}(a,r) = f_U(a+r)/\mu_U$ and f_U is the probability density function associated to

$$F_U(t) = \int_0^t s^{-1}\,dG(s) \Big/ \int_0^\infty s^{-1}\,dG(s). \tag{6.1}$$

The conditional density functions above are explicitly given by

$$f_{A,R|\Delta=1}(a,r) = \frac{1-F_D(r)}{p(a+r)}\,dG(a+r)$$

and

$$f_{A,D|\Delta=0}(a,d) = \frac{f_D(d)}{(1-p)}\int_{a+d\leq z} z^{-1}\,dG(z)$$

for the uncensored and censored subjects, respectively. Here, f_D and F_D are, respectively, the density and distribution functions associated to the residual censoring random variable D, and $p = \mathrm{pr}(\Delta = 1)$ is the proportion of uncensored individuals. The full likelihood of m uncensored and n censored length-biased observations is thus

$$\begin{aligned}\mathcal{L} &= \prod_{i=1}^m \left[\frac{1-F_D(r_i)}{px_i}\,dG(x_i)\right]\prod_{j=1}^n\left[\frac{f_D(d_j)}{1-p}\int_{y_j\leq z} z^{-1}\,dG(z)\right]\\ &\propto \prod_{i=1}^m dG(x_i)\prod_{j=1}^n \int_{y_j\leq z} z^{-1}\,dG(z).\end{aligned}$$

Denoting $G_*(t) = P(A+R\leq t \mid \Delta=1)$ and $F_*(t) = P(A+D\leq t\mid\Delta=0)$ with associated density functions $g_*(t)$ and $f_*(t)$, we may verify that

$$g_*(t) = \frac{g(t)}{pt}\int_0^t [1-F_D(r)]\,dr \quad \text{and} \quad f_*(t) = \frac{f(t)F_D(t)}{1-p},$$

where $f(t)$ is given by (1.1). Defining the operators

$$\mathcal{H}(u)(t) = \int_{0<x\le t} \frac{g_*(x)}{g(x)}\,du(x),$$

$$\mathcal{K}_{m,n}(u)(t) = \int_{0<y\le t} y\biggl(\int_{y\le z} \frac{u(z)}{z^2}\,dz\biggr)\,d\biggl[\biggl(\frac{\hat f(t)}{\hat f(y)} - 1\biggr)\frac{f_*(y)}{f(y)}\biggr]$$

and $\Psi_{m,n} = \hat p \mathcal{H} + (1-\hat p)\mathcal{K}_{m,n}$, Asgharian and Wolfson [3] have derived, under this scenario, the equation $\Psi_{m,n}(U_{m,n}) = W_{m,n}$, where $W_{m,n}$ is obtained from (2.2) by replacing $W_{X,m}$ and $W_{Y,n}$ by the empirical processes $\sqrt{m}(G_m - G_*)$ and $\sqrt{n}(F_n - F_*)$, respectively. Defining the limiting operators

$$\mathcal{K}(u)(t) = \int_{0<y\le t} y\biggl(\int_{y\le z} \frac{u(z)}{z^2}\,dz\biggr)\,d\biggl[\biggl(\frac{f(t)}{f(y)} - 1\biggr)\frac{f_*(y)}{f(y)}\biggr]$$

and $\Psi = p\mathcal{H} + (1-p)\mathcal{K}$, one can show that $\Psi_{m,n}$ converges almost surely to Ψ in operator norm topology, and that Ψ is bounded, linear and has bounded inverse Ψ^{-1} if $p > 0.59$; see [3].

As discussed in the Introduction, when the observation mechanism generates length-biased samples, it is often of prime interest to make inference about F_U and its density function f_U. Substitution of G by $\hat G$ in (6.1) yields $\hat F_U$, an asymptotically efficient estimator of F_U. The asymptotic properties of $Z_{m,n} = \sqrt{k}(\hat F_U - F_U)$ may be studied via its relation to $U_{m,n}$. Indeed, defining $L_t(x) = x^{-1}[\mathbb{I}_{[0,t]}(x) - F_U(t)]$, we may write

$$\hat F_U(s) - F_U(s) = \frac{\int_0^\infty L_s(x)\,d[\hat G(x) - G(x)]}{\int_0^\infty x^{-1}\,d\hat G(x)},$$

from which we have that $Z_{m,n} = \int_0^\infty L_s(x)\,dU_{m,n}(x)/\int_0^\infty x^{-1}\,d\hat G(x)$. Defining the operator $\mathscr{L}(g)(t) = \mu_U^{-1}\int_0^\infty L_t(x)\,dg(x)$, we note that if there exists some $\gamma_0 > 0$ such that $G(\gamma_0) = 0$ (in which case G is said to satisfy assumption γ), the operator $\mathscr{L}$ is bounded. Consequently, Theorems 1–5 hold when making inference about F_U and its density function f_U.

Under the additional assumption that the residual censoring distribution does not have a point-mass at zero, it is possible to provide an explicit distributional result for the empirical density process arising from kernel density estimation. Specifically, we have that the empirical density process $\sqrt{kh_{m,n}}(\hat f_U - f_U)$ is asymptotically Gaussian with mean zero and covariance function σ_{f_U} estimated consistently by

$$\hat\sigma_{f_U}(s,t) = h_{m,n}^{-1}\int\!\!\int \hat\psi_Z(s - uh_{m,n}, t - vh_{m,n})\,dK(u)\,dK(v),$$

where $\hat{\psi}_Z$ is a consistent estimator of the asymptotic covariance function ψ_Z associated to the sequence of processes $Z_{m,n}$. For example, we may take

$$\hat{\psi}_Z(s,t) = \hat{\mu}_U^{-2} \int\int \hat{\psi}_U(x,y)\, d\hat{L}_s(x)\, d\hat{L}_t(y),$$

where $\hat{\mu}_U = \int_0^\infty z^{-1}\, d\hat{G}(z)$, $\hat{L}_u(z) = z^{-1}[\mathbb{I}_{[0,u]}(z) - \hat{F}_U(z)]$ and $\hat{\psi}_U$ is a consistent estimator of the covariance function ψ_U of process U. Since for $s \le t$ we may write $\psi_U(s,t)$ as

$$\begin{aligned} p\bigg\{&\int_0^s [\beta(x)]^2\, dG_*(x) - \bigg[\int_0^s \beta(x)\, dG_*(x) \int_0^t \beta(x)\, dG_*(x)\bigg]\bigg\} \\ &+ (1-p)\int_0^t\int_0^s f(x)f(y)\bigg\{e(x\wedge y) \\ &\qquad\qquad + h(x\wedge y)\bigg[\frac{1}{f(x\vee y)} - \frac{1}{f(x\wedge y)}\bigg] \\ &\qquad\qquad\qquad - h(x)h(y)\bigg\}\, d\zeta(x)\, d\zeta(y), \end{aligned}$$

where we have defined $\zeta(x) = g(x)[pg_*(x)]^{-1}$, $h(x) = \int_0^x F_*(y)\, d[1/f(y)]$ and $e(x) = 2\int_0^x h(y)\, d[1/f(y)]$, consistent estimation of ψ_U is possible by substitution of appropriate empirical counterparts into the above.

Assumption γ imposed on G may seem restrictive, but nonetheless holds in many industrial and medical applications. The case of survival with dementia, studied in [2] and [53], is an example of such. It is possible to relax this requirement by imposing that G and F_D vanish at zero at a super-polynomial rate, that is, by assuming that $G(t)$ and $F_D(t)$ are $o(t^r)$ as $t \to 0$ for each $r > 0$. While preserving all results pertaining to G, this relaxation does not directly preserve those pertaining to F_U. The unboundedness of $\mathscr{L}$ is problematic, although an application of Tikhonov's regularization method may help in circumventing this problem. This has been explored by Carroll, Rooij and Ruymgaart [12], although not from the perspective of strong approximations.

7. Closing remarks. (1) For distributions with a lighter left tail ($\alpha_0 > 2$) and heavier right tail (small β), the rate obtained for the strong approximation of $U_{m,n}$ is close to $k^{-1/4}$ modulo logarithmic terms. It is unclear whether it is possible to achieve better rates; if so, different techniques would necessarily be needed to control the rate of $\mathcal{I}_5$ in Lemma 4, as the best achievable rate for $\mathcal{I}_5$ using approximations by Bernstein polynomials is $k^{-1/4}$. As for assumption (B2) on the bandwidth required to establish Theorem 3, the $k^{-1/4}$ rate in the strong approximation roughly translates into

the bandwidth condition $(\log k)^2/(k^{3/4}h_{m,n}) \to 0$ when we further replace the iterated logarithmic term by a logarithmic term. This is in contrast to $\log k/(kh_{m,n}) \to 0$ obtained in [42], in the case of uncensored observations alone. Likewise, the rate given in Remark 2, after Theorem 4, is roughly $h_{m,n} \sim (\log k)^{2/3}/k^{1/4}$.

(2) The theory presented in this paper requires that $\hat{p} \to p \in (0,1]$. The case $p = 0$ may itself be of interest. On one hand, if $\hat{p} = 0$ for each k, then all observations are multiplicatively censored; this has been studied by Groeneboom [23], among others. On the other hand, if $\hat{p} > 0$ for each k, the methods developed in this paper may be adapted as long as $\hat{p}$ does not vanish too rapidly. Specifically, we may redefine $\mathcal{F}_{m,n} = \hat{p}\mathcal{I} + (1-\hat{p})\mathcal{G}$ and

$$W^0_{m,n}(s) = \sqrt{\hat{p}}B_{X,m}(G(s)) + \sqrt{1-\hat{p}}f(s)\int_{0<y\leq s} B_{Y,n}(F(y))\,d\left[\frac{1}{f(y)}\right].$$

Suppose that $\hat{p}^{-2}$ is $\mathcal{O}(\upsilon_k)$ for some sequence of positive real numbers υ_k tending to infinity. Then the strong approximation holds, with $U^0_{m,n}$ redefined as the Gaussian process $\mathcal{F}^{-1}_{m,n}(W^0_{m,n})$ and the rates being multiplied by $\mathcal{O}(\upsilon_k^2)$. Further, the rate of the global modulus of continuity of $U^0_{m,n}$ is multiplied by $\mathcal{O}(\upsilon_k)$. This allows one to study the case $p = 0$. This extension provides insight into the leap between the square-root asymptotics in the canonical multiplicative censoring setting and the cube-root asymptotics for the Grenander estimator when only censored observations are available.

APPENDIX: PROOFS OF MAIN RESULTS

Proof of Claim 1. If the condition $\sum_{m,n} G(\gamma_{m,n}) < \infty$ is satisfied, it is an immediate consequence of Theorem 1 of Section 10.1 of [41] that $\mathrm{pr}(\min(X_1,\ldots,X_m) \leq \gamma_{m,n} \text{ i.o.}) = 0$. Hence, almost surely, we may find m_0 and $n_0 \in \mathbb{N}$ such that, for each $m \geq m_0$ and $n \geq n_0$, all uncensored observations $x_1,\ldots,x_m$ are no smaller than $\gamma_{m,n}$. We restrict our attention here to such sufficiently large m and n. Define $\delta_i = \mathbb{I}_{\{x_1,\ldots,x_m\}}(t_i)$ for $i = 1,\ldots,k$, and write $r_0 = \min\{i : t_i \geq \gamma_{m,n}\}$. By construction, we must have that $\delta_1 = \cdots = \delta_{r_0-1} = 0$. Define the set

$$\mathscr{D} = \left\{(a_{r_0}, a_{r_0+1},\ldots,a_k) : 0 \leq a_{r_0}, a_{r_0+1},\ldots,a_k \leq 1, \sum_{i=r_0}^{k} a_i = 1, a_k \geq \frac{1}{k}\right\},$$

a bounded, closed and convex subset of $\mathbb{R}^{k-r_0+1}$. For $i = r_0,\ldots,k$, define

$$\begin{aligned}\phi_i(a_{r_0},\ldots,a_k) &= \delta_i\left(\frac{\hat{p}}{m}\right) + \frac{a_i}{t_i}\left(\frac{1-\hat{p}}{n}\right)\sum_{j=1}^{i}\frac{1-\delta_j}{\sum_{q=\max(j,r_0)}^{k} a_q/t_q}\\ &= \frac{1}{k}\left(\delta_i + \frac{a_i}{t_i}\sum_{j=1}^{i}\frac{1-\delta_j}{\sum_{q=\max(j,r_0)}^{k} a_q/t_q}\right)\end{aligned}$$

and $\phi = (\phi_{r_0}, \dots, \phi_k)$. We note that ϕ is continuous on $\mathscr{D}$. We want to show that $\phi(\mathscr{D}) \subseteq \mathscr{D}$. The fact that the image of $\mathscr{D}$ under ϕ_i is contained in $[0,1]$ for $i = r_0, \dots, k$ is clear. That it is contained in $[1/k, 1]$ for $i = k$ is obvious if $\delta_k = 1$. We assume instead that $\delta_k = 0$. Then, defining $\lambda_j = \sum_{q=\max(j,r_0)}^{k-1} a_q/t_q \geq 0$ for $j = 1, \dots, k-1$ and $\lambda_k = a_k/t_k \geq 0$, we observe that

$$\frac{a_k}{t_k} \sum_{j=1}^{k} \frac{1-\delta_j}{\sum_{q=\max(j,r_0)}^{k} a_q/t_q} = \lambda_k \left(\sum_{j=1}^{k-1} \frac{1-\delta_j}{\lambda_j + \lambda_k} + \frac{1}{\lambda_k} \right) \geq 1,$$

from which it follows that the image of $\mathscr{D}$ under ϕ_k is contained in $[1/k, 1]$ if $\delta_k = 0$ as well. Finally, we require the equality $\sum_{i=r_0}^{k} \phi_i(a_{r_0}, \dots, a_k) = 1$ to hold for any $(a_{r_0}, \dots, a_k) \in \mathscr{D}$. This can be verified using that

$$\sum_{i=r_0}^{k} \sum_{j=1}^{i} b_{ij} = \sum_{j=1}^{r_0-1} \sum_{i=r_0}^{k} b_{ij} + \sum_{j=r_0}^{k} \sum_{i=j}^{k} b_{ij}$$

for any array b_{ij}, where under the first sum on the right-hand side, it holds that $\max(j, r_0) = r_0$, while under the second sum, $\max(j, r_0) = j$. We may thus use the Brouwer fixed point theorem (see, e.g., Proposition 2.6 on page 52 and Problem 6.7e on page 254 of [54]) to obtain that there exists some $a^* = (a^*_{r_0}, \dots, a^*_k) \in \mathscr{D}$ such that $\phi(a^*) = a^*$. The distribution function

$$\hat{G}^*(t) = \sum_{i=r_0}^{k} a^*_i \mathbb{I}_{[0,t]}(t_i)$$

is a solution to equation (2.1) with zero mass below $\gamma_{m,n}$. $\square$

PROOF OF THEOREM 1. Using Lemma 1 and the boundedness of $\mathcal{F}^{-1}_{m,n}$, we have for each $t \in [0, \tau - \varepsilon]$ that

$$U_{m,n}(t) = \mathcal{F}^{-1}_{m,n,\varepsilon}(W_{m,n})(t) + \mathcal{O}(\varepsilon\sqrt{\log\log k}) \qquad \text{a.s.}$$

Similarly, using the definition of $U^0_{m,n}$, $W^0_{m,n}$, Lemma 6 and the boundedness of $\mathcal{F}^{-1}$, we have for each $t \in [0, \tau - \varepsilon]$ that

$$U^0_{m,n}(t) = \mathcal{F}^{-1}_{\varepsilon}(W^0_{m,n})(t) + \mathcal{O}(\varepsilon\sqrt{\log k}) \qquad \text{a.s.}$$

The result follows from Lemmas 3, 5 and 6 and the inequality

$$\begin{aligned}
\|U_{m,n} - U^0_{m,n}\|_{[0,\tau-\varepsilon]} &= \|\mathcal{F}^{-1}_{m,n}(W_{m,n}) - \mathcal{F}^{-1}(W^0_{m,n})\|_{[0,\tau-\varepsilon]} \\
&\leq \|\mathcal{F}^{-1}_{m,n,\varepsilon}\| \|W_{m,n} - W^0_{m,n}\|_{[0,\tau-\varepsilon]} \\
&\quad + \|\mathcal{F}^{-1}_{m,n,\varepsilon} - \mathcal{F}^{-1}_{\varepsilon}\| \|W^0_{m,n}\|_{[0,\tau-\varepsilon]} \\
&\quad + \mathcal{O}(\varepsilon\sqrt{\log k}) \qquad \text{a.s.}
\end{aligned}$$

We therefore find that

$$\begin{aligned}\|U_{m,n}-U^0_{m,n}\|_{[0,\tau-\varepsilon]} &\le \mathcal{O}\biggl(\frac{k^{-r(\alpha)}\sqrt{\log k}(\log\log k)^{1/4}}{f(\tau-\varepsilon)}\biggr)\\ &\quad+\mathcal{O}\biggl(\frac{k^{-(\alpha-1)/(2\alpha)}\log k\sqrt{\log\log k}}{f(\tau-\varepsilon)}\biggr)\mathcal{O}(\sqrt{\log k})\\ &\quad+\mathcal{O}(\varepsilon\sqrt{\log k}) \qquad \text{a.s.}\end{aligned}$$

The use of Lemma 3 was justified by the fact that $W^0_{m,n}$ is almost surely continuous. Since (A4) implies that $f(\tau-u)\sim u^\beta$ for u small, the above bound has least order, modulo logarithmic terms, for $\varepsilon=\varepsilon_{m,n}$. $\square$

Proof of Theorem 2. Let $t\in[0,\tau-\eta]$ and $s\in[0,h_{m,n}]$. By definition (3.1), linearity of $\mathcal{F}^{-1}$ and the triangle inequality, we have that

$$|U^0_{m,n}(t+s)-U^0_{m,n}(t)|\le I_m(s,t)+J_n(s,t), \tag{A.1}$$

where we define

$$\begin{aligned}I_m(s,t)&=|\mathcal{F}^{-1}(B_{X,m}\circ G)(t+s)-\mathcal{F}^{-1}(B_{X,m}\circ G)(t)|,\\ J_n(s,t)&=|\mathcal{F}^{-1}(\mathcal{H}_n)(t+s)-\mathcal{F}^{-1}(\mathcal{H}_n)(t)|\end{aligned}$$

and

$$\mathcal{H}_n(t)=f(t)\int_{0<y\le t}B_{Y,n}(F(y))\,d\biggl[\frac{1}{f(y)}\biggr].$$

We first study $I_m(s,t)$. Writing $\varsigma(u)(\cdot)=\int_0^\infty K(\cdot,x)u(x)\,dx$ and noting that $\int_0^\infty \mathcal{A}_0(\cdot,x)u(x)\,dx=\mathcal{G}(u)(\cdot)$ for each u, equations (2.6) and (2.7) imply that $\varsigma(u)\equiv-(1-p)\mathcal{G}(u+p\varsigma(u))/p^2$. It follows from (A5) that $M_1=\|f'\|_{[0,\tau]}<\infty$. We find that

$$\begin{aligned}&|\mathcal{G}(w)(t+s)-\mathcal{G}(w)(s)|\\ &\quad\le|f(t+s)-f(t)|\biggl|\int_{0<y\le t}y\biggl(\int_{y\le z}\frac{w(z)}{z^2}\,dz\biggr)\,d\biggl[\frac{1}{f(y)}\biggr]\biggr|\\ &\qquad+|f(t)|\biggl|\int_{t<y\le t+s}y\biggl(\int_{y\le z}\frac{w(z)}{z^2}\,dz\biggr)\,d\biggl[\frac{1}{f(y)}\biggr]\biggr|\\ &\quad\le|f(t+s)-f(t)|\biggl[\frac{1}{f(t)}-\frac{1}{f(0)}\biggr]\|w\|_{[0,\tau]}\\ &\qquad+|f(t+s)|\biggl[\frac{1}{f(t+s)}-\frac{1}{f(t)}\biggr]\|w\|_{[0,\tau]}\\ &\quad=|f(t+s)-f(t)|\biggl[\frac{f(0)-f(t)}{f(0)f(t)}+\frac{1}{f(t)}\biggr]\|w\|_{[0,\tau]}\end{aligned}$$

$$\leq \frac{2}{f(\tau-\eta)}\|w\|_{[0,\tau]}|f(t+s)-f(t)|$$
$$\leq \frac{2M_1 s}{f(\tau-\eta)}\|w\|_{[0,\tau]}$$

from which it follows, using (2.5), that

$$\begin{aligned}|\varsigma(u)(t+s)-\varsigma(u)(t)| &\leq \frac{2(1-p)M_1 s}{p^2 f(\tau-\eta)}\|u+p\varsigma(u)\|_{[0,\tau]} \\ &\leq \frac{2(1-p)M_1 s}{p^2 f(\tau-\eta)}(2+p\|\mathcal{F}^{-1}\|)\|u\|_{[0,\tau]}.\end{aligned} \tag{A.2}$$

Using (2.5) once more, we then have that

$$\begin{aligned}&\sup_{0\leq t\leq \tau-\eta}\sup_{0\leq s\leq h_{m,n}} I_m(s,t) \\ &\quad\leq p^{-1}\sup_{0\leq t\leq \tau-\eta}\sup_{0\leq s\leq h_{m,n}}|B_{X,m}(G(t+s))-B_{X,m}(G(t))| \\ &\qquad+\sup_{0\leq t\leq \tau-\eta}\sup_{0\leq s\leq h_{m,n}}|\varsigma(B_{X,m}\circ G)(t+s)-\varsigma(B_{X,m}\circ G)(t)|.\end{aligned}$$

Using (A5), we may show, as in [31] and [41], that

$$\sup_{0\leq x\leq a_\tau}\sup_{0\leq y\leq M_0 h_{m,n}}|\mathcal{W}_{X,m}(x+y)-\mathcal{W}_{X,m}(x)| = \mathcal{O}(\sqrt{h_{m,n}\log(1/h_{m,n})})$$

almost surely, where $a_\tau = G(\tau-\eta)$, $M_0 = \|g\|_{[0,\tau]}$, and $\mathcal{W}_{X,m}$ is the Wiener process associated with $B_{X,m}$; see Lemma 1.4.1 of [15]. Hence, by an application of the MVT, $B_{X,m}\circ G$ has modulus of continuity

$$\mathcal{O}(\sqrt{h_{m,n}\log(1/h_{m,n})})$$

as well. In view of (A.2) and the fact that $\|B_{X,m}\circ G\|_{[0,\tau]}$ is $\mathcal{O}(\sqrt{\log m})$ almost surely, we have that

$$\sup_{0\leq t\leq \tau-\eta}\sup_{0\leq s\leq h_{m,n}}|\varsigma(B_{X,m}\circ G)(t+s)-\varsigma(B_{X,m}\circ G)(t)| = \mathcal{O}(\sqrt{\log m}\,h_{m,n})$$

almost surely. It follows from the discussion above then that

$$\sup_{0\leq t\leq \tau-\eta}\sup_{0\leq s\leq h_{m,n}} I_m(s,t) = \mathcal{O}(\sqrt{h_{m,n}\log(1/h_{m,n})}) \qquad \text{a.s.} \tag{A.3}$$

We now turn to $J_n(s,t)$. Defining

$$J_n'(s,t) = |f(t+s)-f(t)|\int_{0<y\leq t}|B_{Y,n}(F(y))|\,d\left[\frac{1}{f(y)}\right]$$

and

$$J_n''(s,t) = |f(t+s)| \int_{t<y\le t+s} |B_{Y,n}(F(y))|\, d\left[\frac{1}{f(y)}\right],$$

we notice that $|\mathcal{H}_n(t+s) - \mathcal{H}_n(t)| \le J_n'(s,t) + J_n''(s,t)$. Using the MVT, we have that

$$J_n'(s,t) \le \frac{M_1 s}{f(\tau-\eta)} \sup_{0\le y\le 1} |B_{Y,n}(y)|$$

and

$$J_n''(s,t) \le \frac{f(t) - f(t+s)}{f(t)} \sup_{0\le y\le 1} |B_{Y,n}(y)| \le \frac{M_1 s}{f(t)} \sup_{0\le y\le 1} |B_{Y,n}(y)|,$$

so that $\sup_{0\le t\le \tau-\eta} \sup_{0\le s\le h_{m,n}} J_n'(s,t)$, $\sup_{0\le t\le \tau-\eta} \sup_{0\le s\le h_{m,n}} J_n''(s,t)$ and consequently $\sup_{0\le t\le \tau-\eta} \sup_{0\le s\le h_{m,n}} |\mathcal{H}_n(t+s) - \mathcal{H}_n(t)|$ are $\mathcal{O}(\sqrt{\log n}\, h_{m,n})$ almost surely. Further, using (A.2), we have that

$$\begin{aligned}
&\sup_{0\le t\le \tau-\eta} \sup_{0\le s\le h_{m,n}} |\varsigma(\mathcal{H}_n)(t+s) - \varsigma(\mathcal{H}_n)(t)| \\
&\quad \le \frac{2(1-p)M_1}{p^2 f(\tau-\eta)} (2 + p\cdot \|\mathcal{F}^{-1}\|) \|\mathcal{H}_n\|_{[0,\tau]} h_{m,n} = \mathcal{O}(\sqrt{\log n}\, h_{m,n}) \qquad \text{a.s.}
\end{aligned}$$

so that $\sup_{0\le t\le \tau-\eta} \sup_{0\le s\le h_{m,n}} J_n(s,t) = \mathcal{O}(\sqrt{\log n}\, h_{m,n})$ almost surely using (2.5). The theorem follows in view of this last result, (A.1) and (A.3). □

Proof of Theorem 3. By the continuity (and hence uniform continuity) of g on $[0,\tau]$, the dominated convergence theorem may be used to show that

$$\lim_{k\to\infty} \sup_{0\le s\le \tau-\eta} |g_{m,n}(s) - g(s)| = 0. \tag{A.4}$$

The theorem follows immediately from Lemma 7 and the triangle inequality. □

Proof of Theorem 4. By Theorem 1 and integration by parts, for any $t \in [0, \tau-\eta]$, we may write that

$$\begin{aligned}
\hat{g}_{m,n}(t) - g(t) &= [\hat{g}_{m,n}(t) - g_{m,n}(t)] + [g_{m,n}(t) - g(t)] \\
&= \frac{1}{\sqrt{k}} \int_0^\infty U_{m,n}^0(s)\, d\psi_{m,n}(t,s) \\
&\quad + \mathcal{O}\left(\frac{V_{m,n}\varepsilon_{m,n}(\log k)^{3/2}\sqrt{\log\log k}}{\sqrt{k}} + \delta_{m,n} \right) \qquad \text{a.s.,}
\end{aligned}$$

where $\delta_{m,n} = \sup_{0\le t\le \tau-\eta} |g_{m,n}(t) - g(t)|$. The result follows from [37], which shows that $\delta_{m,n} = \mathcal{O}(h_{m,n}^2)$ and $V_{m,n} = \mathcal{O}(1/h_{m,n})$. □

Proof of Theorem 5. Since g is twice continuously differentiable on $[0,\tau-\eta]$, we may write that $g_{m,n}(s)-g(s)=h_{m,n}^2\sigma^2 g''(s)/2+o(h_{m,n}^2)$ uniformly in $s\in[0,\tau-\eta]$. Combining this expansion with (S.1) in the proof of Lemma 7 (see supplementary material [1]) yields

$$\begin{aligned}\hat{g}_{m,n}(s)-g(s)&=\left(\frac{h_{m,n}^2\sigma^2}{2}\right)g''(s)+\frac{\Upsilon_{m,n}(s,h_{m,n})}{\sqrt{k}h_{m,n}}\\&\quad+\mathcal{O}\left(\frac{\varepsilon_{m,n}(\log k)^{3/2}\sqrt{\log\log k}}{\sqrt{k}h_{m,n}}\right)+o(h_{m,n}^2)\qquad\text{a.s.}\end{aligned}$$

uniformly in $s\in[0,\tau-\eta]$, where $\Upsilon_{m,n}(s,h)=\int_{-1}^1 U_{m,n}^0(s-uh)\,dK(u)$. In view of (2.5) and the proof of Theorem 2, we find that

$$\Upsilon_{m,n}(s,h_{m,n})=p^{-1/2}\int_{-1}^1 B_{X,m}(G(s-uh_{m,n}))\,dK(u)+\mathcal{O}(\sqrt{\log k}h_{m,n})\quad\text{a.s.}$$

Further, using (B3) we may show, for $\alpha\ge 2$, that

$$\frac{\varepsilon_{m,n}(\log k)^{3/2}\sqrt{\log\log k}}{\sqrt{k}h_{m,n}}=o(h_{m,n}^2)$$

and therefore that

$$\hat{g}_{m,n}(s)-g(s)=\left(\frac{h_{m,n}^2\sigma^2}{2}\right)g''(s)+\frac{\int_{-1}^1 B_{X,m}(G(s-uh))\,dK(u)}{\sqrt{pk}h_{m,n}}+o(h_{m,n}^2)$$

almost surely. It then follows that $\mathcal{E}_{m,n}(0,\tau-\eta)$ may be written as

$$\begin{aligned}&\frac{h_{m,n}^4\sigma^4}{4}\int_0^{\tau-\eta}\{g''(s)\}^2\,ds+\frac{{}_\eta P_{m,n}(h_{m,n})}{pkh_{m,n}^2}+\frac{\sigma^2 h_{m,n}{}_\eta Q_{m,n}(h_{m,n})}{\sqrt{pk}}\\&\quad+o(h_{m,n}^2)\left\{o(h_{m,n}^2)+h_{m,n}^2\sigma^2\int_0^{\tau-\eta}g''(s)\,ds+\frac{2{}_\eta R_{m,n}(h_{m,n})}{\sqrt{pk}h_{m,n}}\right\}\qquad\text{a.s.,}\end{aligned}$$

where we have defined

$${}_\eta P_{m,n}(h)=\int_0^{\tau-\eta}\left[\int_{-1}^1 B_{X,m}(G(s-uh))\,dK(u)\right]^2 ds,$$

$${}_\eta Q_{m,n}(h)=\int_0^{\tau-\eta}g''(s)\left[\int_{-1}^1 B_{X,m}(G(s-uh))\,dK(u)\right]ds$$

and

$${}_\eta R_{m,n}(h)=\int_0^{\tau-\eta}\left[\int_{-1}^1 B_{X,m}(G(s-uh))\,dK(u)\right]ds.$$

It follows from [24] that ${}_\eta P_{m,n}(h) = h_{m,n}\nu^2 + o_p(h_{m,n})$, while ${}_\eta Q_{m,n}(h)$ and ${}_\eta R_{m,n}(h)$ are both $o_p(\sqrt{h_{m,n}})$. We therefore obtain that $\mathcal{E}_{m,n}(0, \tau - \eta)$ may be expressed as

$$\frac{h_{m,n}^4 \sigma^4}{4} \int_0^{\tau-\eta} [g''(s)]^2 \, ds + \frac{\nu^2}{pkh_{m,n}} + o_p\left(\frac{1}{kh_{m,n}} + h_{m,n}\sqrt{\frac{h_{m,n}}{k}} + h_{m,n}^4\right).$$

The result follows upon noticing that a term of order $o_p(h_{m,n}^{3/2}/\sqrt{k})$ is dominated by any term of order $o_p(h_{m,n}^4)$. $\square$

Acknowledgments. The authors thank the current and previous Editors, Professors Cai and Silverman, the anonymous Associate Editor and the referees for their exceptionally insightful and constructive comments.

SUPPLEMENTARY MATERIAL

Additional technical details: Proof of lemmas
(DOI: 10.1214/11-AOS954SUPP; .pdf). The proof of each lemma in the paper is provided in the supplementary material.

REFERENCES

[1] Asgharian, M., Carone, M. and Fakoor, V. (2012). Supplement to "Large-sample study of the kernel density estimators under multiplicative censoring." DOI:10.1214/11-AOS954SUPP.

[2] Asgharian, M., M'Lan, C. E. and Wolfson, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Amer. Statist. Assoc.* **97** 201–209. MR1947280

[3] Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann. Statist.* **33** 2109–2131. MR2211081

[4] Asgharian, M., Wolfson, D. B. and Zhang, X. (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. *Stat. Med.* **25** 1751–1767. MR2227351

[5] Bergeron, P.-J., Asgharian, M. and Wolfson, D. B. (2008). Covariate bias induced by length-biased sampling of failure times. *J. Amer. Statist. Assoc.* **103** 737–742. MR2524006

[6] Bickel, P. J., Klaassen, A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore.

[7] Bickel, P. J. and Ritov, J. (1991). Large sample theory of estimation in biased sampling regression models. I. *Ann. Statist.* **19** 797–816. MR1105845

[8] Bikel, P. D. and Ritov, I. (1994). Efficient estimation using both direct and indirect observations. *Theory Probab. Appl.* **38** 194–213.

[9] Blum, J. R. and Susarla, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data. In *Multivariate Analysis, V (Proc. Fifth Internat. Sympos., Univ. Pittsburgh, Pittsburgh, PA, 1978)* 213–222. North-Holland, Amsterdam. MR0566340

[10] Burke, M. D., Csörgő, S. and Horváth, L. (1981). Strong approximations of some biometric estimates under random censorship. *Probab. Theory Related Fields* **56** 87–112.

[11] Burke, M. D., Csörgő, S. and Horváth, L. (1988). A correction to and improvement of: "Strong approximations of some biometric estimates under random censorship." *Probab. Theory Related Fields* **79** 51–57. MR0952993

[12] Carroll, R. J., van Rooij, A. C. M. and Ruymgaart, F. H. (1991). Theoretical aspects of ill-posed problems in statistics. *Acta Appl. Math.* **24** 113–140. MR1131834

[13] Cox, D. R. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling* (N. L. Johnson and H. Smith, eds.). Wiley, New York.

[14] Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London. MR0751780

[15] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York. MR0666546

[16] Csörgő, S. and Hall, P. (1984). The Komlós–Major–Tusnády approximations and their applications. *Austral. J. Statist.* **26** 189–218. MR0766619

[17] del Barrio, E., Deheuvels, P. and van de Geer, S. (2007). *Lectures on Empirical Processes: Theory and Statistical Applications*. Eur. Math. Soc., Zürich. MR2284824

[18] Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics* **6** 13–25.

[19] Földes, A., Rejtő, L. and Winter, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data. II. Estimation of density and failure rate. *Period. Math. Hungar.* **12** 15–29. MR0607625

[20] Gilbert, P. B. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *Ann. Statist.* **28** 151–194. MR1762907

[21] Gill, R. D., Vardi, Y. and Wellner, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16** 1069–1112. MR0959189

[22] Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415

[23] Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, CA, 1983)* 539–555. Wadsworth, Belmont, CA. MR0822052

[24] Hall, P. (1982). Limit theorems for stochastic measures of the accuracy of density estimators. *Stochastic Process. Appl.* **13** 11–25. MR0662801

[25] Hasminskii, R. Z. and Ibragimov, I. A. (1983). On asymptotic efficiency in the presence of an infinite-dimensional nuisance parameter. In *Probability Theory and Mathematical Statistics (Tbilisi, 1982). Lecture Notes in Math.* **1021** 195–229. Springer, Berlin. MR0735986

[26] Huang, J. and Wellner, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065

[27] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412

[28] Kvam, P. (2008). Length bias in the measurements of carbon nanotubes. *Technometrics* **50** 462–467. MR2655646

[29] Marron, J. S. and Padgett, W. J. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *Ann. Statist.* **15** 1520–1535. MR0913571

[30] Mason, D. M. (2007). Some observations on the KMT dyadic scheme. *J. Statist. Plann. Inference* **137** 895–906. MR2301724

[31] Mason, D. M., Shorack, G. R. and Wellner, J. A. (1983). Strong limit theorems for oscillation moduli of the uniform empirical process. *Probab. Theory Related Fields* **65** 83–97.

[32] Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data. *Ann. Statist.* **14** 766–773. MR0840530

[33] Nadaraja, È. A. (1965). On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.* **10** 199–203.

[34] Neyman, J. (1955). Statistics: Servant of all science. *Science* **122** 401–406.

[35] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959

[36] Parthasarathy, K. R. (2005). *Probability Measures on Metric Spaces*. AMS Chelsea Publishing, Providence, RI. MR2169627

[37] Révész, P. (1972). On empirical density function. *Period. Math. Hungar.* **2** 85–110. MR0368287

[38] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837. MR0079873

[39] Schuster, E. F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40** 1187–1195. MR0247723

[40] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York. MR1191168

[41] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York. MR0838963

[42] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184. MR0471166

[43] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London. MR0848134

[44] Steele, J. M. (1978). Invalidity of average squared error criterion in density estimation. *Canad. J. Statist.* **6** 193–200. MR0532858

[45] van der Vaart, A. (1994). Maximum likelihood estimation with partially censored data. *Ann. Statist.* **22** 1896–1916. MR1329174

[46] Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge Univ. Press, New York.

[47] Van Ryzin, J. (1969). On strong consistency of density estimates. *Ann. Math. Statist.* **40** 1765–1772. MR0258172

[48] Vardi, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10** 616–620. MR0653536

[49] Vardi, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–205. MR0773161

[50] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76** 751–761. MR1041420

[51] Vardi, Y. and Zhang, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20** 1022–1039. MR1165604

[52] Wicksell, S. D. (1925). The corpuscle problem. *Biometrika* **17** 84–99.

[53] Wolfson, C., Wolfson, D. B., Asgharian, M., M'Lan, C. E., Ostbye, T., Rockwood, K. and Hogan, D. B. et al. (2001). A reevaluation of the duration of survival after the onset of dementia. *New England Journal of Medicine* **344** 1111.

[54] Zeidler, E. (1985). *Nonlinear Functional Analysis and Its Applications, Part I: Fixed-Point Theorems*. Springer, New York.

[55] Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56** 601–614. MR0258224

[56] Zhang, B. (1998). A note on the integrated square errors of kernel density estimators under random censorship. *Stochastic Process. Appl.* **75** 225–234. MR1632201

M. Asgharian
Department of Mathematics and Statistics
McGill University
Burnside Hall, Room 1237
805 rue Sherbrooke Ouest
Montréal, QC, H3A 2K6
Canada
E-mail: masoud@math.mcgill.ca

M. Carone
Division of Biostatistics
University of California, Berkeley
109 Haviland Hall
Berkeley, California 94720-7358
USA
E-mail: mcarone@berkeley.edu

V. Fakoor
Department of Statistics
Ferdowsi University of Mashhad
Azadi Square
Mashhad, Razavi Khorasan 1159-91775
Iran
E-mail: fakoor@math.um.ac.ir